

PAT-NO: JP410134052A

DOCUMENT-IDENTIFIER: JP 10134052 A

TITLE: MACHINE TRANSLATION METHOD/DEVICE

PUBN-DATE: May 22, 1998

INVENTOR-INFORMATION:

NAME
KOBAYAKAWA, KENSHUKURU
TAZOE, HIDEKAZU
ISHIKURA, JUN

ASSIGNEE-INFORMATION:

| NAME | COUNTRY |
|------------------------------------|---------|
| INTERNATL BUSINESS MACH CORP <IBM> | N/A |

APPL-NO: JP08272755

APPL-DATE: October 15, 1996

INT-CL (IPC): G06F017/28, G06F013/00 , G06F017/21

ABSTRACT:

PROBLEM TO BE SOLVED: To suitably translate a text in a WWW(world wide web) page which is down-loaded from a WWW server on an internet into a translated text.

SOLUTION: When a user inputs a URL(uniform resource locator) character string on a WWW browser 150 screen, the designated WWW page is down-loaded via the internet 70. At that time, the URL character string which is similar (namely, maximum likelihood) to the inputted URL character sting is retrieved on a data base. Then, a translation engine 120 translates a foreign language text in the WWW page which is down-loaded into a native language text based on selected translation environment. The native language text being a translation processing result is transferred to the WWW browser 150 and is displayed on the browser 150 screen.

COPYRIGHT: (C)1998,JP

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平10-134052

(43)公開日 平成10年(1998)5月22日

(51)Int.Cl.⁶

G 06 F 17/28
13/00
17/21

識別記号

3 5 1

F I

G 06 F 15/20
13/00
15/20
15/38

5 9 2 A
3 5 1 G
5 9 6 A
Z

審査請求 未請求 請求項の数13 OL (全 17 頁)

(21)出願番号

特願平8-272755

(22)出願日

平成8年(1996)10月15日

(71)出願人 390009531

インターナショナル・ビジネス・マシーンズ・コーポレーション
INTERNATIONAL BUSINESSES MACHINES CORPORATION
アメリカ合衆国10504、ニューヨーク州
アーモンク (番地なし)

(72)発明者 小早川 健 シュクリ

神奈川県大和市下鶴間1623番地14 日本アイ・ピー・エム株式会社 大和事業所内

(74)代理人 弁理士 合田 漢 (外2名)

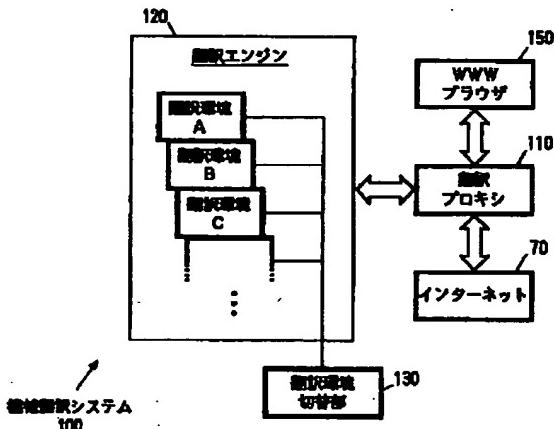
最終頁に続く

(54)【発明の名称】 機械翻訳方法及び装置

(57)【要約】 (修正有)

【課題】 インターネット上のWWWサーバからダウンロードしたWWWページ中の原文テキストを訳文テキストに、好適に翻訳する。

【解決手段】 WWWブラウザ画面上でユーザがURL文字列を入力すると、指定されたWWWページがインターネット経由でダウンロードされる。このとき、入力されたURL文字列に類似した（すなわち最尤な）URL文字列をデータベース上で検索する。次いで、翻訳エンジンは、選択された翻訳環境を基に、ダウンロードされたWWWページ中の外国語テキストを母国語テキストに翻訳処理する。翻訳処理結果である母国語テキストは、WWWブラウザに渡されてブラウザ画面上に表示される。



【特許請求の範囲】

【請求項1】複数の翻訳環境を有する翻訳システムを用いて、ユーザが指定した情報リソースに格納されたファイル中のテキストを翻訳するための機械翻訳方法において、(a)情報リソースへのアクセス手段と情報リソース名を記述した組合せ文字列を解釈する段階と、(b)解釈結果に従って前記複数の翻訳環境のうちの1つを選択する段階と、(c)前記組合せ文字列が示した情報リソースに格納されたファイル中のテキストを、前記選択された翻訳環境を用いて翻訳する段階と、を具備することを特徴とする機械翻訳方法。

【請求項2】複数の翻訳環境を有する翻訳システムを用いて、ユーザが指定した情報リソースに格納されたファイル中のテキストを翻訳するための機械翻訳方法において、(a)情報リソースへのアクセス手段と情報リソース名を記述した組合せ文字列をユーザーが入力する段階と、(b)前記組合せ文字列が指定する情報リソースに格納されたファイルを、前記組合せ文字列が指定するアクセス手段に従ってダウンロードする段階と、(c)前記組合せ文字列を解釈する段階と、(d)解釈結果に従って前記複数の翻訳環境のうちの1つを選択する段階と、(e)前記ダウンロードされたファイル中のテキストを、前記選択された翻訳環境を用いて翻訳する段階と、を具備することを特徴とする機械翻訳方法。

【請求項3】複数の翻訳環境を有する翻訳システムを用いて、ユーザが指定した情報リソースに格納されたファイル中のテキストを翻訳するための機械翻訳方法において、(a)情報リソースへのアクセス手段と情報リソース名を記述した組合せ文字列をユーザーが入力する段階と、(b)前記組合せ文字列が指定する情報リソースに格納されたファイルを、前記組合せ文字列が指定するアクセス手段に従ってダウンロードする段階と、(c)ダウンロードされたファイルをファイル中の原文テキストとともにディスプレイ上に表示する段階と、(d)前記組合せ文字列を解釈する段階と、(e)解釈結果に従って前記複数の翻訳環境のうちの1つを選択する段階と、(f)前記原テキストを、前記選択された翻訳環境を用いて翻訳する段階と、(g)翻訳処理後の訳文テキストを前記原文テキストとともにディスプレイ上に表示する段階と、を具備することを特徴とする機械翻訳方法。

【請求項4】複数の翻訳環境を有する翻訳システムを用いて、情報リソースへのアクセス手段と情報リソース名を記述した組合せ文字列の入力に応じて指定された情報リソースに格納されたファイルをダウンロードするとともにファイル中のテキストを翻訳するための機械翻訳方法において、(a)情報リソースに係る組合せ文字列と翻訳環境との組合せをデータベースとして記録する段階と、(b)前記データベース中で、新たに入力された組合せ文字列に最尤な組合せ文字列を検索する段階と、

のテキストを、前記検索された組合せ文字列に該当する翻訳環境を用いて翻訳する段階と、を具備することを特徴とする機械翻訳方法。

【請求項5】複数の翻訳環境を有する翻訳システムを用いて、情報リソースへのアクセス手段と情報リソース名を記述した組合せ文字列の入力に応じて指定された情報リソースに格納されたファイルをダウンロードするとともにファイル中のテキストを翻訳するための機械翻訳方法において、(a)情報リソースに係る組合せ文字列と翻訳環境との組合せをデータベースとして記録する段階と、(b)前記データベース中で、新たに入力された組合せ文字列に最尤な組合せ文字列を検索する段階と、(c)情報リソースからダウンロードされたファイル中のテキストを、前記検索された組合せ文字列に係る翻訳環境を用いて翻訳する段階と、(d)翻訳結果が適切でなければ他の翻訳環境を用いて前記テキストを再度翻訳処理する段階と、を具備することを特徴とする機械翻訳方法。

【請求項6】複数の翻訳環境を有する翻訳システムを用いて、情報リソースへのアクセス手段と情報リソース名を記述した組合せ文字列の入力に応じて指定された情報リソースに格納されたファイルをダウンロードするとともにファイル中のテキストを翻訳するための機械翻訳方法において、(a)情報リソースに係る組合せ文字列と翻訳環境との組合せをデータベースとして記録する段階と、(b)前記データベース中で、新たに入力された組合せ文字列に最尤な組合せ文字列を検索する段階と、(c)情報リソースからダウンロードされたファイル中のテキストを、前記検索された組合せ文字列に係る翻訳環境を用いて翻訳する段階と、(d)翻訳結果が適切でなければ他の翻訳環境を用いて前記テキストを再度翻訳処理する段階と、(e)翻訳結果が適切な場合には、新たな組合せ文字列と翻訳環境との関係を前記データベース中に追加記録する段階と、を具備することを特徴とする機械翻訳方法。

【請求項7】情報リソースにアクセスするためのプロトコル名、情報リソースを格納するサーバ名（及びサーバ内のファイル名）の組合せからなるURL（Uniform Resource Locator）文字列の入力に応じて、指定されたサーバ内のファイルを指定されたプロトコルに従ってインターネット経由でダウンロードし、複数の翻訳環境を用いてダウンロードされたファイル中のテキストを翻訳するための機械翻訳方法において、(a)入力されたURL文字列を解釈する段階と、(b)解釈結果に従って前記複数の翻訳環境のうちの1つを選択する段階と、(c)ダウンロードされたファイル中のテキストを、前記選択された翻訳環境を用いて翻訳する段階と、を具備することを特徴とする機械翻訳方法。

【請求項8】情報リソースにアクセスするためのプロトコル名、情報リソースを格納するサーバ名（及びサーバ

内のファイル名)の組合せからなるURL(Uniform Resource Locator)文字列の入力に応じて、指定されたサーバ内のファイルを指定されたプロトコルに従ってインターネット経由でダウンロードし、複数の翻訳環境を用いてダウンロードされたファイル中のテキストを翻訳するための機械翻訳方法において、(a)ダウンロードされたファイルをファイル中の原文テキストとともにディスプレイ上に表示する段階と、(b)前記入力されたURL文字列を解釈する段階と、(c)解釈結果に従って前記複数の翻訳環境のうちの1つを選択する段階と、

(d)前記原テキストを、前記選択された翻訳環境を用いて翻訳する段階と、(e)翻訳処理後の訳文テキストを前記原文テキストとともにディスプレイ上に表示する段階と、を具備することを特徴とする機械翻訳方法。

【請求項9】情報リソースにアクセスするためのプロトコル名、情報リソースを格納するサーバ名(及びサーバ内のファイル名)の組合せからなるURL(Uniform Resource Locator)文字列の入力に応じて、指定されたサーバ内のファイルを指定されたプロトコルに従ってインターネット経由でダウンロードし、複数の翻訳環境を用いてダウンロードされたファイル中のテキストを翻訳するための機械翻訳方法において、(a)URL文字列と翻訳環境との組合せをデータベースとして記録する段階と、(b)前記データベース中で、新たに入力されたURL文字列に最尤なURL文字列を検索する段階と、(c)ダウンロードされたされたファイル中のテキストを、前記検索されたURL文字列に該当する翻訳環境を用いて翻訳する段階と、を具備することを特徴とする機械翻訳方法。

【請求項10】情報リソースにアクセスするためのプロトコル名、情報リソースを格納するサーバ名(及びサーバ内のファイル名)の組合せからなるURL(Uniform Resource Locator)文字列の入力に応じて、指定されたサーバ内のファイルを指定されたプロトコルに従ってインターネット経由でダウンロードし、複数の翻訳環境を用いてダウンロードされたファイル中のテキストを翻訳するための機械翻訳方法において、(a)URL文字列と翻訳環境との組合せをデータベースとして記録する段階と、(b)前記データベース中で、新たに入力されたURL文字列に最尤なURL文字列を検索する段階と、(c)ダウンロードされたファイル中のテキストを、前記検索されたURL文字列に該当する翻訳環境を用いて翻訳する段階と、(d)翻訳結果が適切でなければ他の翻訳環境を用いて前記テキストを再度翻訳処理する段階と、を具備することを特徴とする機械翻訳方法。

【請求項11】情報リソースにアクセスするためのプロトコル名、情報リソースを格納するサーバ名(及びサーバ内のファイル名)の組合せからなるURL(Uniform Resource Locator)文字列の入力に応じて、指定されたサーバ内のファイルを指定されたプロトコルに従ってイ

ンターネット経由でダウンロードし、複数の翻訳環境を用いてダウンロードされたファイル中のテキストを翻訳するための機械翻訳方法において、(a)URL文字列と翻訳環境との組合せをデータベースとして記録する段階と、(b)前記データベース中で、新たに入力されたURL文字列に最尤なURL文字列を検索する段階と、(c)ダウンロードされたファイル中のテキストを、前記検索されたURL文字列に係る翻訳環境を用いて翻訳する段階と、(d)翻訳結果が適切でなければ他の翻訳環境を用いて前記テキストを再度翻訳処理する段階と、

(e)翻訳結果が適切な場合には、新たなURL文字列と翻訳環境との関係を前記データベース中に追加記録する段階と、を具備することを特徴とする機械翻訳方法。

【請求項12】インターネット経由でWWWサーバからダウンロードされたファイル中のテキストを、複数の翻訳環境を用いて翻訳するための機械翻訳装置において、(a)URL文字列と翻訳環境との組合せを記述したデータベースを記録する第1の手段と、(b)ユーザがURLを入力するための第2の手段と、(c)入力されたURLで指定されたWWWサーバのファイルをダウンロードするための第3の手段と、(d)入力されたURL文字列に最尤なURL文字列を前記第1の手段の中から検索するための第4の手段と、(e)検索されたURL文字列に係る翻訳環境を最適なものとして選択するための第5の手段と、(f)ダウンロードしたWWWサーバのファイル中のテキストを、前記選択された翻訳環境を用いて翻訳する第6の手段と、(g)翻訳結果を画面表示するための第7の手段と、を具備することを特徴とする機械翻訳装置。

【請求項13】インターネット経由でWWWサーバからダウンロードされたファイル中のテキストを翻訳するためのコンピュータ・プログラムを格納したコンピュータ可読記憶媒体であって、前記コンピュータ・プログラムは、(a)複数の翻訳環境と、(b)URL文字列と翻訳環境との組合せを記述したデータベースを記録したデータベースと、(c)ユーザが入力したURLが指定するWWWサーバのファイルをダウンロードするための第1のコンピュータ可読プログラム・コード手段と、(d)入力されたURL文字列に最尤なURL文字列を前記データベースの中から検索する第2のコンピュータ可読プログラム・コード手段と、(e)検索されたURL文字列に係る翻訳環境を最適なものとして選択する第3のコンピュータ可読プログラム・コード手段と、

(f)ダウンロードしたWWWサーバのファイル中のテキストを、前記選択された翻訳環境を用いて翻訳する第4のコンピュータ可読プログラム・コード手段と、(g)翻訳結果を画面表示するための第4のコンピュータ可読プログラム・コード手段と、を具備することを特徴とするコンピュータ可読記憶媒体。

50 【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、例えば英語等の外国語テキストを日本語（あるいは韓国語や中国語など）の母国語テキストに翻訳するための機械翻訳方法及び装置に係り、特に、複数の翻訳環境を備えた機械翻訳方法及び装置に関する。更に詳しくは、本発明は、インターネット上のWWWサーバによって提供されたWWWページ中の外国語テキストを母国語テキストに翻訳するための機械翻訳方法に係り、特に、処理対象の外国語テキストに適した翻訳環境を自動選択して翻訳処理する機械翻訳方法及び装置に関する。なお、ここで言う「翻訳環境」とは、翻訳処理に利用される辞書データベースや文法処理アルゴリズム（例えば文節の設定や助動詞訳の設定、文体の指定など）などを指す。

【0002】

【従来の技術】

インターネット (Internet) : 従来より、一般公衆回線等の通信回線を用いて遠隔地のコンピュータ同士をつなぐ「テレコンピューティング・システム」の研究開発が行われてきた。例えば日本国内では、1985年の電気通信事業法の施行に端を発してパソコン通信が盛んになってきた。

【0003】最近のテレコンピューティングの主要な動向として、インターネット (Internet) の利用が挙げられよう。ここで、インターネットとは、1970年代に米国の全米科学財団 (NSF) が出資して設立されたネットワーク "NSF net" を母体として相互接続された世界中のネットワークの総称である。インターネットは、そもそもは学術研究の情報交換を初期目的として世界54ヶ国の大学や研究機関等結ぶための国際的なネットワークであるが、各大学・研究機関に設置されたサーバ（主にUNIXワークステーション）が自動的に相互接続を繰り返した結果、全世界を網羅する巨大なネットワークに成長した（現在のインターネットは、全世界に散在する数十万を越えるサーバで構成される！）。遠隔地間の各サーバ・マシンは、"TCP/IP (Transmission Control Protocol/Internet Protocol)" と呼ばれる通信プロトコルに従って相互接続されており、各マシン間でデータ転送やBBS (Bulletin Board System: 電子掲示板) 上でのコミュニケーションなどが可能である。

【0004】従来の（すなわち原初的な）インターネットは、UNIXベースのコンピュータ同士の接続、というイメージが強かったが、1994年になって一般公開され、インターネット接続サービスを提供するプロバイダが出現するに至って、一般企業や個人にも普及が加速してきた。例えば日本の主要パソコン通信サービス（"PC-VAN"や"Niifty Serve"など）でもインターネット接続サービスが提供されている。

【0005】インターネットは、運営母体やポリシーの

異なるネットワークの集合体であり、個々のネットワークを統括する機関はあってもインターネット全体を包括的に管理する機関は存在しない。すなわち、世界中に散在する各サーバ（例えばデータベース・サーバやファイル・サーバ）は、ネットワーク内のユーザによって局所的に管理されているに過ぎない。したがって、インターネットの各ユーザは、インターネットに接続したサーバを介することにより、誰もが情報提供者となり得、また、インターネット上のあらゆる情報にシームレスにアクセスすることができる。この点は、特定のホスト（すなわち情報サーバ）にしかアクセスできないパソコン通信サービスとは大いに異なっている。なお、特定企業のネットワーク同士を接続してワイドエリアに広げたインターネットワーキングと区別するために、英語では "the Internet" と固有名詞化して表記するのが一般的となっている。

【0006】ワールド・ワイド・ウェブ (World Wide Web: WWW) : 今日のインターネットの普及を加速したのは「ワールド・ワイド・ウェブ (WWW)」である、といつても過言ではなかろう。WWWとは、ネットワーク（特にインターネット）上にハイパーテキスト//を構築し、あらゆる情報にシームレスにアクセスすることを目的とした広域情報システム（インターネット・アプリケーション）のことであり、もともとは1989年に欧州素粒子物理学研究所（通称CERN）で開発された。WWWによれば、ユーザ（すなわちWWWクライアント）は、画面上で強調表示されたワード、フレーズ、画像等をマウスでクリックすることによって、サーバからサーバへ（若しくはデータベースからデータベースへ）、目的の情報へ辿り着くことが可能である。

【0007】「世界中に広がった蜘蛛の巣」の名の通り、現在では非常に多数のWWWサーバがインターネット上に存在している。各WWWサーバのローカル・ディスクには、HTML (Hypertext markup Language) と呼ばれる言語で記述されたファイルからなるWWWページが多数蓄積されている。

【0008】HTMLファイルは、テキスト・ファイルの他、画像、音声、ビデオ等の各種形式のファイルを含んでいる。HTMLは、SGML (Standard Generalized Markup Language) を基に策定されており、現在のバージョンはWWWの標準化団体 "W3 Consortium" が1996年にリリースした "HTML 3. x" である。HTMLに準拠したファイルは、基本的にはテキスト形式のファイルで構成されており、"<"と">"で区切られた「予約語（タグ）」を使って、テキストの整形や画像ファイル・音声ファイルの在処、リンク先等が記述されている。

【0009】WWWユーザー（WWWクライアント）は、WWWサーバのWWWページに格納されたファイルを、HTTP (Hypertext Transfer Protocol) という通信

プロトコルに従って読み出すことができる（HTTPは、TCP/IPの上位プロトコルに該当する）。いわゆる「WWWブラウザ」は、HTMLファイルの内容を解釈してユーザー・マシンの画面上に表示する、すなわちインターネット・ツアーガイドとして作用するコンピュータ・プログラムである。WWWブラウザの代表例は、米IBM社の"Web Explorer"（"Web Explorer"は米IBM社の商標）、米Netscape Communications社の"Netscape Navigator"や米Microsoft社の"Internet Explorer"等である。

【0010】HTMLファイル中に指定された画像ファイル等の在処は、HTMLファイルそのものを保存しているWWWサーバと別のWWWサーバであっても構わない。WWWブラウザは、HTMLファイルの解釈を行う他、画像ファイル等の読み出しも行って、テキスト中に画像等が埋め込まれた複合テキストをWWWクライアント・マシンのディスプレイ画面上に表示するようになっている。また、HTMLファイル中には「このテキスト（若しくは画像）がマウスでクリックされたら、どのサーバのどのファイルを読み出す」という如きコマンドが埋め込まれており、ユーザは表示画面上のテキストや画面の一部分をマウスでクリック操作するだけで別のHTMLファイルを読み出すことができる。クリック操作により新たに読み出されるHTMLファイルは、元のHTMLファイルと同一のWWWサーバにあっても別のWWWサーバにあっても構わない。

【0011】HTMLファイル中に次のHTMLファイルの読み出しを指示したコマンドを埋め込むことを、当業界では「リンクを張る」とも言う。リンク先の指定には、URL (Uniform Resource Locator) と呼ぶ表記方法が用いられる。URLは、インターネット上の情報リソース（すなわちインターネット上のサーバが蓄積しているファイル）を識別するための標準的なアドレッシング技術であり、IETF (Internet Engineering Task Force) /***/ がまとめた公式文書群RFC (Request for Comments) の1つであるRFC1738でその仕様が規定されている。

【0012】URLは、情報リソースにアクセスするためのプロトコル名（スキーム名）、情報リソースを格納するサーバ名、及びサーバ内のファイル名という3要素を記述した文字列で構成される。指定できるプロトコルには、HTTP (http) の他、NNTP (news)、FTP (ftp: file transfer protocol)、Telnet (telnet)、ローカル・ディスクへのアクセス (file) 等が挙げられる。また、主なサーバ名には、"www.ibm.com"や"www.yahoo.com"等が挙げられる。例えば、"www.yahoo.com"というサーバのWWWサービスで提供するファイル"Arts"のURLは、"http://www.yahoo.com/Arts/"と表記されよう。URLの各構成要素の間は、通常、"/"で区切られている。換言すれば、URLは、WWWペ

ージのアドレスなのである。（なお、ここで言う「WWWページ」とは、WWWサーバが提供する画面の総称のことであり、「ホームページ」とも言う。）

【0013】なお、URLを構成する文字列のうち、"/"で区切られた各文字列のことを、本明細書中では以下「パーシャルURL文字列」と呼ぶこととする。

【0014】WWWページのための翻訳プログラム：前述したように、WWWに依れば、インターネット・ユーザはデスクに居ながらにして世界中のWWWサーバにア

10 クセスし、取得したWWWページをブラウザ画面上で眺めることができる。

【0015】WWWページ中のテキストの多くは英語で記述されており、英語圏以外（例えば日本、中国、韓国等）のインターネット・ユーザ（若しくは英語に精通しないユーザ）にとっては解説が不便であり、折角の広域情報システムを充分に活用することができない。このため、最近では、機械翻訳ソフト（若しくは翻訳支援ソフト）の開発・普及化が再燃している。これら機械翻訳ソフトの中には、インターネット上のテキストすなわちW

20 WWページの翻訳に特化した製品もある。例えば、インターネット経由で取得したデータを、WWWブラウザのプロキシ（Proxy）機能/***/を利用して機械翻訳ソフトが中継し、翻訳文が再びWWWブラウザに渡されてブラウザ画面に表示される、という具合である。このようなインターネットとの融和性を重視した機械翻訳ソフトの一例は、日本アイ・ビー・エム（株）が市販する「インターネット翻訳の王様 Ver1.x」である。

【0016】機械翻訳とは、要するにコンピュータ・システムの処理能力を活用して、例えば英語等の外国語テキストを日本語などの母国語テキストに翻訳する（若しくはその逆方向に翻訳する）技術のことであり、そもそも、人間が有する（若しくは有すると推定される）言語に関する莫大な知識をモデル化したデータベースを構築し、翻訳エンジンがこのデータベースを参照しながら行うものである。

【0017】ところが、原文テキストの意図を忠実に反映させながら精度よく機械翻訳することは、概して困難である。これは、自然言語の持つ曖昧さにも大いに依存する。曖昧さの最たる要因の1つとして、多義語すなわち2以上の意味を持つ単語が存在し、且つ原文テキスト中に頻出するということが挙げられよう。多義語が出現するたびに翻訳エンジンは訳語選択しなければならず、訳語選択にはある程度の誤差すなわち誤訳を伴うことは言うまでもないであろう。また、曖昧さの他の要因として、文法上のバリエーションが挙げられよう。例えば英文の場合、主語・述語等の関係からSVO, SVOO, SVOCなど幾つかの基本構文型に分類されるが、関係詞 ("that"や"which", "why"など) を省略する等により、文型や文法の把握が困難となる場合が多い。

40 【0018】機械翻訳における翻訳精度を向上させるた

50

めに、複数の翻訳環境を備えるという手法が従来より広く採り入れられている。ここで言う「翻訳環境」とは、翻訳処理に利用される辞書データベースや文法処理アルゴリズム（例えば文節の設定や助動詞訳の設定、文体の指定など）などのことを指すと把握されたい。原文テキストの内容が属するジャンル、分野、用途などと翻訳環境との間に相関関係があることは既に一般に認められている。例えば芸術、スポーツ、教育、科学技術など、各分野に応じて頻出される単語、単語の意味、独特の言い回しなどは互に異なり、インターネット辞書、アート辞書、スポーツ辞書などジャンル別に専用の翻訳環境が既に利用可能となっている。つまり、原文テキストの内容に最適な翻訳環境を用いることによって、機械翻訳時ににおける不必要的解析処理が軽減され、翻訳精度が向上するという訳である。

【0019】複数の翻訳環境の中から最適なものを選択するための最も簡易且つ原始的な手法は、ユーザ自らが原文テキストの内容から判断してマニュアル選択する、ということであろう。この場合、ユーザは、機械翻訳の開始前に原文テキストの内容をある程度は理解しておく必要がある。しかしながら、ユーザが母国語でない原文テキストの内容を一目で理解することは容易ではなく、多くの時間を費やしかねない。この結果、翻訳システムの存在自体が意味を持たないものとなる。特にインターネット・ユーザの多くは、莫大なWWWページの中から有意義な情報を一目で取捨選択するために機械翻訳の助けを借りるから、原文テキストを眺める時間をしていては機械翻訳の意味そのものがなくなる。

【0020】これに対し、機械翻訳システムに翻訳環境の選択処理を委ねる、という先行技術もある。例えば特開平7-191999号公報、特開平6-332946号公報、特開平6-318229号公報、特開平6-60117号公報、及び特開昭61-173060号等には、翻訳システムが翻訳環境を自動選択するタイプの翻訳システムについて開示されている。この種の翻訳システムは、ユーザ自らが翻訳環境を選択するという煩雑な手続を取り除き、作業効率を向上させてくれるであろう。しかしながら、これら既存技術の殆どは、原文テキストの意図に適合する翻訳環境を選択するために、翻訳システムが原文テキストの内容の解析を予め行うようになっている。原文テキストの解析というプレプロセスを含むということは、処理に相当の時間を要し、しかも翻訳システムを重たくしてしまう。言い換えれば、これら既存技術は構築・活用の両面で重たいシステムであった訳である。

【0021】《注釈》

* : ハイパーテキストとは、文字や絵、表などをオブジェクトとして扱い、さらに各オブジェクト間で様々な関係付けをしてアクセスできるようにしたソフトウェアである。

** : IETFとは、インターネット上の技術、サービスの議論、提案等を行う機関のことである。

*** : プロキシ機能とは、インターネットからの通信データを代理で受け取り、直接外部にアクセスできないホストに中継する機能のことである。

【0022】

【発明が解決しようとする課題】本発明の目的は、複数の翻訳環境を備えた翻訳システムを用いて、外国語テキスト（例えば英語テキスト）を母国語テキスト（例えば日本語、あるいは韓国語や中国語など）に翻訳するための、優れた機械翻訳方法及び装置を提供することにある。

【0023】本発明の更なる目的は、外国語テキストを母国語テキストに翻訳する際に、処理対象の外国語テキストに適した翻訳環境を自動選択して翻訳処理する、優れた機械翻訳方法及び装置を提供することにある。

【0024】本発明の更なる目的は、複数の翻訳環境を備えた翻訳システムを用いて、インターネット上のWWWサーバによって提供されたWWWページ中の外国語テキストを母国語テキストに好適に翻訳することができる、優れた機械翻訳方法及び装置を提供することにある。

【0025】

【課題を解決するための手段】本発明は、上記課題を参考してなされたものであり、その第1の側面は、複数の翻訳環境を有する翻訳システムを用いて、ユーザが指定した情報リソースに格納されたファイル中のテキストを翻訳するための機械翻訳方法において、（a）情報リソースへのアクセス手段と情報リソース名を記述した組合せ文字列を解釈する段階と、（b）解釈結果に従って前記複数の翻訳環境のうちの1つを選択する段階と、（c）前記組合せ文字列が示した情報リソースに格納されたファイル中のテキストを、前記選択された翻訳環境を用いて翻訳する段階と、を具備することを特徴とする機械翻訳方法である。

【0026】また、本発明の第2の側面は、複数の翻訳環境を有する翻訳システムを用いて、ユーザが指定した情報リソースに格納されたファイル中のテキストを翻訳するための機械翻訳方法において、（a）情報リソースへのアクセス手段と情報リソース名を記述した組合せ文字列をユーザーが入力する段階と、（b）前記組合せ文字列が指定する情報リソースに格納されたファイルを、前記組合せ文字列が指定するアクセス手段に従ってダウンロードする段階と、（c）前記組合せ文字列を解釈する段階と、（d）解釈結果に従って前記複数の翻訳環境のうちの1つを選択する段階と、（e）前記ダウンロードされたファイル中のテキストを、前記選択された翻訳環境を用いて翻訳する段階と、を具備することを特徴とする機械翻訳方法である。

【0027】また、本発明の第3の側面は、複数の翻訳

11

環境を有する翻訳システムを用いて、ユーザが指定した情報リソースに格納されたファイル中のテキストを翻訳するための機械翻訳方法において、(a)情報リソースへのアクセス手段と情報リソース名を記述した組合せ文字列をユーザが入力する段階と、(b)前記組合せ文字列が指定する情報リソースに格納されたファイルを、前記組合せ文字列が指定するアクセス手段に従ってダウンロードする段階と、(c)ダウンロードされたファイルをファイル中の原文テキストとともにディスプレイ上に表示する段階と、(d)前記組合せ文字列を解釈する段階と、(e)解釈結果に従って前記複数の翻訳環境のうちの1つを選択する段階と、(f)前記原テキストを、前記選択された翻訳環境を用いて翻訳する段階と、(g)翻訳処理後の訳文テキストを前記原文テキストとともにディスプレイ上に表示する段階と、を具備することを特徴とする機械翻訳方法である。

【0028】また、本発明の第4の側面は、複数の翻訳環境を有する翻訳システムを用いて、情報リソースへのアクセス手段と情報リソース名を記述した組合せ文字列の入力に応じて指定された情報リソースに格納されたファイルをダウンロードするとともにファイル中のテキストを翻訳するための機械翻訳方法において、(a)情報リソースに係る組合せ文字列と翻訳環境との組合せをデータベースとして記録する段階と、(b)前記データベース中で、新たに入力された組合せ文字列に最尤な組合せ文字列を検索する段階と、(c)情報リソースからダウンロードされたファイル中のテキストを、前記検索された組合せ文字列に該当する翻訳環境を用いて翻訳する段階と、を具備することを特徴とする機械翻訳方法である。

【0029】また、本発明の第5の側面は、複数の翻訳環境を有する翻訳システムを用いて、情報リソースへのアクセス手段と情報リソース名を記述した組合せ文字列の入力に応じて指定された情報リソースに格納されたファイルをダウンロードするとともにファイル中のテキストを翻訳するための機械翻訳方法において、(a)情報リソースに係る組合せ文字列と翻訳環境との組合せをデータベースとして記録する段階と、(b)前記データベース中で、新たに入力された組合せ文字列に最尤な組合せ文字列を検索する段階と、(c)情報リソースからダウンロードされたファイル中のテキストを、前記検索された組合せ文字列に係る翻訳環境を用いて翻訳する段階と、(d)翻訳結果が適切でなければ他の翻訳環境を用いて前記テキストを再度翻訳処理する段階と、を具備することを特徴とする機械翻訳方法である。

【0030】また、本発明の第6の側面は、複数の翻訳環境を有する翻訳システムを用いて、情報リソースへのアクセス手段と情報リソース名を記述した組合せ文字列の入力に応じて指定された情報リソースに格納されたファイルをダウンロードするとともにファイル中のテキス

10

20

30

40

50

12

トを翻訳するための機械翻訳方法において、(a)情報リソースに係る組合せ文字列と翻訳環境との組合せをデータベースとして記録する段階と、(b)前記データベース中で、新たに入力された組合せ文字列に最尤な組合せ文字列を検索する段階と、(c)情報リソースからダウンロードされたファイル中のテキストを、前記検索された組合せ文字列に係る翻訳環境を用いて翻訳する段階と、(d)翻訳結果が適切でなければ他の翻訳環境を用いて前記テキストを再度翻訳処理する段階と、(e)翻訳結果が適切な場合には、新たな組合せ文字列と翻訳環境との関係を前記データベース中に追加記録する段階と、を具備することを特徴とする機械翻訳方法である。【0031】また、本発明の第7の側面は、情報リソースにアクセスするためのプロトコル名、情報リソースを格納するサーバ名(及びサーバ内のファイル名)の組合せからなるURL(Uniform Resource Locator)文字列の入力に応じて、指定されたサーバ内のファイルを指定されたプロトコルに従ってインターネット経由でダウンロードし、複数の翻訳環境を用いてダウンロードされたファイル中のテキストを翻訳するための機械翻訳方法において、(a)入力されたURL文字列を解釈する段階と、(b)解釈結果に従って前記複数の翻訳環境のうちの1つを選択する段階と、(c)ダウンロードされたファイル中のテキストを、前記選択された翻訳環境を用いて翻訳する段階と、を具備することを特徴とする機械翻訳方法である。

【0032】また、本発明の第8の側面は、情報リソースにアクセスするためのプロトコル名、情報リソースを格納するサーバ名(及びサーバ内のファイル名)の組合せからなるURL(Uniform Resource Locator)文字列の入力に応じて、指定されたサーバ内のファイルを指定されたプロトコルに従ってインターネット経由でダウンロードし、複数の翻訳環境を用いてダウンロードされたファイル中のテキストを翻訳するための機械翻訳方法において、(a)ダウンロードされたファイルをファイル中の原文テキストとともにディスプレイ上に表示する段階と、(b)前記入力されたURL文字列を解釈する段階と、(c)解釈結果に従って前記複数の翻訳環境のうちの1つを選択する段階と、(d)前記原テキストを、前記選択された翻訳環境を用いて翻訳する段階と、

(e)翻訳処理後の訳文テキストを前記原文テキストとともにディスプレイ上に表示する段階と、を具備することを特徴とする機械翻訳方法である。

【0033】また、本発明の第9の側面は、情報リソースにアクセスするためのプロトコル名、情報リソースを格納するサーバ名(及びサーバ内のファイル名)の組合せからなるURL(Uniform Resource Locator)文字列の入力に応じて、指定されたサーバ内のファイルを指定されたプロトコルに従ってインターネット経由でダウンロードし、複数の翻訳環境を用いてダウンロードされた

13

ファイル中のテキストを翻訳するための機械翻訳方法において、(a) URL文字列と翻訳環境との組合せをデータベースとして記録する段階と、(b) 前記データベース中で、新たに入力されたURL文字列に最尤なURL文字列を検索する段階と、(c) ダウンロードされたされたファイル中のテキストを、前記検索されたURL文字列に該当する翻訳環境を用いて翻訳する段階と、を具備することを特徴とする機械翻訳方法である。

【0034】また、本発明の第10の側面は、情報リソースにアクセスするためのプロトコル名、情報リソースを格納するサーバ名（及びサーバ内のファイル名）の組合せからなるURL（Uniform Resource Locator）文字列の入力に応じて、指定されたサーバ内のファイルを指定されたプロトコルに従ってインターネット経由でダウンロードし、複数の翻訳環境を用いてダウンロードされたファイル中のテキストを翻訳するための機械翻訳方法において、(a) URL文字列と翻訳環境との組合せをデータベースとして記録する段階と、(b) 前記データベース中で、新たに入力されたURL文字列に最尤なURL文字列を検索する段階と、(c) ダウンロードされたファイル中のテキストを、前記検索されたURL文字列に該当する翻訳環境を用いて翻訳する段階と、(d) 翻訳結果が適切でなければ他の翻訳環境を用いて前記テキストを再度翻訳処理する段階と、を具備することを特徴とする機械翻訳方法である。

【0035】また、本発明の第11の側面は、情報リソースにアクセスするためのプロトコル名、情報リソースを格納するサーバ名（及びサーバ内のファイル名）の組合せからなるURL（Uniform Resource Locator）文字列の入力に応じて、指定されたサーバ内のファイルを指定されたプロトコルに従ってインターネット経由でダウンロードし、複数の翻訳環境を用いてダウンロードされたファイル中のテキストを翻訳するための機械翻訳方法において、(a) URL文字列と翻訳環境との組合せをデータベースとして記録する段階と、(b) 前記データベース中で、新たに入力されたURL文字列に最尤なURL文字列を検索する段階と、(c) ダウンロードされたファイル中のテキストを、前記検索されたURL文字列に係る翻訳環境を用いて翻訳する段階と、(d) 翻訳結果が適切でなければ他の翻訳環境を用いて前記テキストを再度翻訳処理する段階と、(e) 翻訳結果が適切な場合には、新たなURL文字列と翻訳環境との関係を前記データベース中に追加記録する段階と、を具備することを特徴とする機械翻訳方法である。

【0036】また、本発明の第12の側面は、インターネット経由でWWWサーバからダウンロードされたファ

14

イル中のテキストを、複数の翻訳環境を用いて翻訳するための機械翻訳装置において、(a) URL文字列と翻訳環境との組合せを記述したデータベースを記録する第1の手段と、(b) ユーザがURLを入力するための第2の手段と、(c) 入力されたURLで指定されたWWWサーバのファイルをダウンロードするための第3の手段と、(d) 入力されたURL文字列に最尤なURL文字列を前記第1の手段の中から検索するための第4の手段と、(e) 検索されたURL文字列に係る翻訳環境を最適なものとして選択するための第5の手段と、(f) ダウンロードしたWWWサーバのファイル中のテキストを、前記選択された翻訳環境を用いて翻訳する第6の手段と、(g) 翻訳結果を画面表示するための第7の手段と、を具備することを特徴とする機械翻訳装置である。

【0037】なお、ここで言う「翻訳環境」とは、翻訳処理に利用される辞書データベースや文法処理アルゴリズム（例えば文節の設定や助動詞訳の設定、文体の指定など）などのことを指すと把握されたい。

【0038】

20 【作用】本発明は、主にWWWページの機械翻訳に向けられたものであり、WWWページのアドレスであるURLがプロトコル名、サーバ名、ファイル名という各要素を記述した文字列（すなわち「パーシャルURL文字列」）の組合せで構成される、という点に着目してなされたものである。

【0039】本発明に係る機械翻訳方法及び装置は、コンピュータ・プロセスという形態で実装可能であり、例えばWWWサーバからWWWページを受信しブラウザ画面に表示するという一連のブラウジング動作に連動して実行される。

30 【0040】該コンピュータ・プロセスは、外国語テキストを母国語テキストに翻訳処理する（またはその逆方向に翻訳処理する）ための翻訳エンジンの他に、翻訳環境A（インターネット辞書）、翻訳環境B（アート辞書）、翻訳環境C（スポーツ辞書）等のような複数の翻訳環境を備えている。また、本発明に係る機械翻訳方法は、予めURL文字列に関するデータベースを蓄積している。このデータベースは、下表1に示すような、比較的頻繁にアクセスされるWWWページに関するURL文字列と翻訳環境との対応テーブルという形態であってもよい。同テーブルで各URL文字列と対応付けられた翻訳環境は、過去の翻訳処理結果等により最適な環境と看做されたものである。

40 【0041】

【表1】

15

16

| URL文字列 | 翻訳環境 |
|---------------------------------------|------------------|
| www.ibm.com/ | 翻訳環境A（インターネット辞書） |
| www.yahoo.com/ | 翻訳環境A（インターネット辞書） |
| www.yahoo.com/Arts/ | 翻訳環境B（アート辞書） |
| www.yahoo.com/Arts/Recreation/Sports/ | 翻訳環境C（スポーツ辞書） |

【0042】表1では、URL名“www.ibm.com/”及び“www.yahoo.com/”に対しては翻訳環境A（インターネット辞書）が、“www.yahoo.com/Arts/”に対しては翻訳環境B（アート辞書）が、“www.yahoo.com/Arts/Recreation/Sports/”に対しては翻訳環境C（スポーツ辞書）が、夫々最適の翻訳環境として記録されている。該データベースは、システムの設計者がユーザーの一般的な傾向を基に予め作成していくてもよい。また、各ユーザーのアクセス状況を学習してデータベースを更に蓄積するようにしてもよい。

【0043】WWWブラウザ画面上でユーザーがURL文字列を入力すると、WWWブラウザ（若しくはWWWブラウザのプロキシ）は、インターネット経由で、URL文字列で指定されたWWWページからファイルをダウンロードする。一方、本発明に係るコンピュータ・プロセスは、入力されたURL文字列を解釈する。ここで言うURL文字列の解釈とは、入力されたURL文字列に最も類似した（すなわち最尤な）URL文字列をデータベース上で検索することを意味する。より具体的には、URL文字列をバーチャルURL文字列の組合せとして捉え、バーチャルURL文字列の組合せが最も一致するURL文字列をデータベース上から見つけ出す訳である。例えばURL入力がデータベースに既存の“www.ibm.com/”であれば、これに該当する翻訳環境A（インターネット辞書）が自動選択される。また、URL入力がデータベースに未登録の“www.yahoo.com/Arts/Architecture/”であれば、バーチャルURL文字列の一一致度が最も高い“www.yahoo.com/Arts/”に該当する翻訳環境B（アート辞書）が自動選択される。なお、本発明に係るコンピュータ・プロセスは、WWWページのダウンロード作業の完了後、あるいはダウンロード作業に並行して、URL文字列の解釈を行ってもよい。または、本発明に係るコンピュータ・プロセスは、WWWブラウザのプロキシ機能を包含して、WWWページのダウンロード作業を代行するという実装形態でもよい。

【0044】次いで、翻訳エンジンは、選択された翻訳環境を基に、ダウンロードされたファイル中の外国語テキストを母国語テキストに翻訳処理する。翻訳処理結果である母国語テキストは、例えばWWWブラウザに渡されてブラウザ画面上に表示される。このとき、元の外国語テキストからなる原文ウィンドウと、母国語テキスト*50

10*からなる訳文ウィンドウを一画面上に同時表示（対訳表示）するようにしてもよい。

【0045】翻訳結果がブラウザ画面に表示された時点で、ユーザーは訳文の内容が適当でないと判断すれば、他の翻訳環境に切り替えて再度機械翻訳するようにしてもよい。また、データベースに未登録のURL文字列について最適な翻訳環境が見つかったときに、データベースに追加登録するようにしてもよい。

【0046】要するに、本発明は、インターネット上のWWWサーバによって提供されたWWWページを機械翻訳する際に、WWWページのアドレスであるURLの文字列を基にして翻訳環境を自動的に切り替える、という機械翻訳方法及び装置である。本発明によれば最適な翻訳環境を選択するので、翻訳精度は自ずと向上する。また、翻訳環境の切替え・データベースの登録等の処理を、URL文字列という比較的少容量のデータのみに基づいて行うので、処理が軽く、且つシステムの構築も容易である。

【0047】なお、機械翻訳の際に複数の翻訳環境の中から原文テキストの意図に適合する翻訳環境を選択する、という類の技術自体は本出願前に既に存在している（前述の公開特許公報参照）。但し、これら既存技術の殆ど全ては、URL文字列ではなく、原テキストの内容を解析した結果を参照して翻訳環境を選択するようになっている。言い換れば、これら既存技術は原文テキストを解析するという処理を含んでおり、構築・活用の両面で重たいシステムであり、本発明とは本質的に相違している。また、これら既存技術は、そもそも、WWWページの翻訳には向けていない。別の見方をすれば、本発明とこれら構成が相違する既存技術とを共存させることも可能である。

【0048】本発明のさらに他の目的、特徴や利点は、後述する本発明の実施例や添付する図面に基づくより詳細な説明によって明らかになるであろう。

【0049】

【発明の実施の形態】以下、図面を参照しながら本発明の実施例を詳解する。

【0050】A. ハードウェア環境

図1には、本発明を実現するのに適したコンピュータ・システム10のハードウェア構成を模式的に示している。以下、各部について説明する。

17

【0051】メイン・コントローラであるCPU11は、オペレーティング・システム(OS)の制御下で、各種プログラムを実行するようになっている。CPU11は、例えば米Intel社製の“Pentium/1xxMHz”でよい。また、OSは、例えば米IBM社の“OS/2”(“OS/2”は米IBM社の商標)や米Microsoft社の“Windows 95”でよい。

【0052】CPU11は、自身の外部ピンに直結したプロセッサ・バス12、ローカル・バスとしてのPCI(Peripheral Component Interconnect)バス16、及び、入出力バスとしてのISA(Industry Standard Architecture)バス18という3階層のバスを介して各ハードウェア・ブロックと相互接続している。

【0053】プロセッサ・バス12とPCIバス16とは、ブリッジ回路(ホスト-PCIブリッジ)13によって連絡されている。本実施例のブリッジ回路13は、メイン・メモリ14へのアクセス動作を制御するメモリ・コントローラや、両バス12、16間の速度差を吸収するためのデータ・バッファなどを含んだ構成となっている。

【0054】メイン・メモリ14は、実行プログラムの読み込み領域として、あるいは実行プログラムの作業領域として用いられ、一般には複数のDRAM(ダイナミックRAM)チップで構成される。また、L2-キャッシュ15は、頻繁にアクセスされるコード/データを一時格納して、CPU11がメイン・メモリ14にアクセスする時間を吸収するために設けられた高速メモリであり、一般にはSRAM(ステティックRAM)チップが用いられる。また、ROM17は、電源投入時のテスト・プログラム(POST)や、システム10内の各ハードウェアを操作するためのコード群を恒久的に格納するため不揮発性メモリである。

【0055】PCIバス16は、比較的高速なデータ転送が可能なバス(バス幅32ビット、動作周波数33MHz、最大データ転送速度132KBps)であり、ビデオ・コントローラ20やカードバス・コントローラ(図示しない)のような比較的高速で動作する周辺デバイス類が接続される。

【0056】ビデオ・コントローラ20は、CPU11からの描画命令を実際に処理するための専用コントローラであり、処理した描画情報を画面バッファ(VRAM)21に一旦書き込むとともに、VRAM21から描画情報を読み出してディスプレイ22にビデオ出力するようになっている。

【0057】PCIバス16とISAバス18とは、ブリッジ回路(PCI-ISAブリッジ)19によって相互接続されている。本実施例のブリッジ回路19は、DMAコントローラやプログラマブル割込みコントローラ(PIC)、プログラマブル・インターバル・タイマ(PIT)を含んだ構成となっている。また、本実施例

18

のブリッジ回路19は、IDE(Integrated Drive Electronics)に準拠した外部記憶装置類を接続するためのIDEコネクタも備えている。IDEコネクタには、例えばIDEハード・ディスク・ドライブ(HDD)23やIDE CD-ROMドライブ24が接続される。

【0058】なお、上述した2つのブリッジ回路13及び19はPCIによって作成された仕様に準拠するものであり、一般には単一のチップセットの形態で提供される。チップセットの一例は、米Intel社が市販する“Triton”である。

【0059】ISAバス18は、PCIバスに比しデータ転送速度が低く(バス幅16ビット、最大データ転送速度4MBps)、キーボード/マウス・コントローラ(KMC)25やI/Oコントローラ25、オーディオ・コントローラ34のような、比較的低速駆動するタイプの周辺デバイス類を接続するのに用いられる。

【0060】KMC25は、キーボード26からの入力コードや、マウス27からの指示座標値を、コンピュータ・データとしてバス18上に取り込むための専用コントローラである。

【0061】I/Oコントローラ28は、フロッピー・ディスク・ドライブ(FDD)29の駆動や、パラレル・ポート30又はシリアル・ポート32を介したパラレル的若しくはシリアル的なデータ転送動作を制御するための専用コントローラである。一般に、パラレル・ポート30にはプリンタ31が、シリアル・ポート32にはモデム33が、夫々接続される。

【0062】モデム33は、アナログ的な一般公衆回線50を介してデータ交換を行うための機器であり、コンピュータ・システム10内のデジタル信号をアナログ信号に変調して回線50上に伝送したり、逆に回線50経由で受信したアナログ信号をデジタル信号に復調してコンピュータ・システム10内に取り込む機能を持っている。本実施例のコンピュータ・システム10は、公衆回線50経由でサーバ・マシン60(より厳密にはサーバ・マシン60のモデル(図示しない))にダイヤルアップ接続することが可能である。ここで言う「サーバ・マシン」とは、インターネット接続サービスを提供する「サービス・プロバイダ」が所有するサーバ・マシンの

40 1つ(例えばシステム10のユーザの最寄のサーバ・マシン)である。サーバ・マシン60は、一般には、インターネット70に常設されており、自身にダイヤルアップしたコンピュータ・システム10をインターネット70にIP接続する。なお、サービス・プロバイダの例は、日本アイ・ビー・エム(株)の「IBMインターネット接続サービス」や、ベッコアメ・インターネットの「BEKKOAME/INTERNET」などである。

【0063】オーディオ・コントローラ34は、オーディオ信号の入出力処理を行うための専用コントローラであり、マイク35から入力されたオーディオ信号をコン

19

ピュータ・データとして取り込んだり、オーディオ・データをDA変換してスピーカ36からオーディオ出力するようになっている。

【0064】なお、現在市販されているいわゆるパーソナル・コンピュータは、図1に示したコンピュータ・システム10として充分に機能を発揮するであろう。なお、コンピュータ・システム10を構成するためには、図1に示した以外にも多くの電気回路等が必要である。但し、これらは当業者には周知であり、また、本発明の要旨を構成するものではないので、本明細書中では省略している。

【0065】B. 機械翻訳システム

図2には、本発明によって提供される機械翻訳システム100のシステム構成を模式的に示している。該機械翻訳システムは、実際には、図1に示すようなコンピュータ・システム10上で実行されるアプリケーション・プログラムの1つという形態で実現される。

【0066】該機械翻訳システム100は、翻訳プロキシ110と、翻訳エンジン120と、翻訳環境切替部130とで構成され、これら各部とWWWブラウザ150との協働的動作により、機械翻訳プロセスが好適に実現されるようになっている。

【0067】WWWブラウザ150は、インターネット・ツアーガイドとして作用するコンピュータ・プログラムであり、コンピュータ・システム10のユーザに対してURL(Uniform Resource Locator)すなわち所望のWWWページのアドレスを入力するためのユーザ・インターフェースを提供するとともに、WWWページに格納されたHTMLファイルの内容を解釈して、ディスプレイ22上に表示するようになっている。本実施例のWWWブラウザ150は、プロキシ機能を備えたタイプである。すなわち、WWWブラウザ150は、WWWページ上のファイルのダウンロードを翻訳プロキシ110に依頼するとともに、翻訳後のテキスト・ファイルを翻訳プロキシ110から受け取って、画像ファイル(例えばGIFファイル)等とともにディスプレイ22上に表示するようになっている。WWWブラウザ150の表示の態様として、例えば「原文のみ」、「訳文のみ」、「対訳表示」などの各表示モードをサポートしている。ブラウザ画面上で切替可能である。また、ブラウザ画面はマルチウィンドウ中の1つのウィンドウでもよい(後述)。

【0068】翻訳プロキシ110は、WWWブラウザ150からの要求に従って、URLで指定されたWWWページからファイルをインターネット70経由でダウンロードするとともに、ダウンロードしたファイル中のテキストを翻訳エンジン120に受け渡し、さらに、翻訳エンジン120から渡された翻訳処理後の訳文テキストをWWWブラウザ150に受け渡すようになっている。なお、「プロキシ」とは、直接外部にアクセスできないホ

20

ストの代理で、インターネットからの通信データを受け取るなどの中継機能の総称である。

【0069】翻訳エンジン120は、機械翻訳処理、すなわち原文(例えば英語等の外国語テキスト)を訳文(例えば日本語等の母国語テキスト)に翻訳する本体であり、翻訳プロキシ110から受け取った原文テキストを自身が備えた翻訳環境を利用しながら翻訳処理し、処理結果である訳文テキストを翻訳プロキシ110に渡すようになっている。本実施例の翻訳エンジン120は、

10 翻訳環境A、翻訳環境B、翻訳環境C…など複数の翻訳環境を備えている。ここで言う「翻訳環境」とは、翻訳処理に利用される辞書データベースや文法処理アルゴリズム(例えば文節の設定や助動詞訳の設定、文体の指定など)などを指す。原文テキストの内容が属するジャンル、分野、用途などと翻訳環境との間に相関関係があることは、既に一般に認められている通りである。本実施例では、例えば、インターネット辞書、ビジネス辞書、政治辞書、エンターテイメント辞書、アート辞書、スポーツ辞書などのように、ジャンル毎に専用の辞書データベースを有している。

【0070】翻訳環境切替部130は、翻訳エンジン120が使用すべき翻訳環境を適宜切り替えるためのものである。本実施例に係る翻訳環境切替部130は、WWWページのアドレスであるURLがプロトコル名、サーバ名、ファイル名という各要素を記述した文字列(すなわち「バーチャルURL文字列」)の組合せで構成されるという点に着目してデザインされたものであり、WWWブラウザ150に対して入力されたURL文字列に基づいて最適な翻訳環境を自動選択するようになっている。

30 【0071】図3には、翻訳環境切替部130の構成を模式的に示している。同図に示すように、本実施例の翻訳環境切替部130は、URL入力部131と、翻訳環境出力部132と、URLデータベース133と、URL照合部134と、基本操作部135とに細分化される。

【0072】URLデータベース133は、比較的頻繁にアクセスされるWWWページのURL文字列と、その各々に対して最適と看做されている翻訳環境との対応関係を記述したテーブルである。

【0073】図4及び図5には、URLデータベース133をWWWブラウザ画面上に表示した様子を示している(但し、図5は図4の画面の続き)。両図に示すように、比較的頻繁にアクセスされる各WWWサーバ/WWWページのURL文字列は、それぞれに最適な翻訳環境と関連付けてデータベース133上に登録されている。例えば、"www.ibm.com/"や"www.yahoo.com/"は「インターネット一般」が、また、"cnn.com/"は「ビジネス」が、夫々最適な翻訳環境として登録されている。本実施例のURLデータベース133には、各URL文字

50

21

列が不規則的に（すなわち登録順等に従うだけで、アルファベットやパーシャルURL文字列の並びや個数などのいずれの秩序にも依らないで）、記述されている。

【0074】URL文字列と最適な翻訳環境の対応関係として特徴的な点の1つは、URL文字列の示すWWWサーバ名が同じであっても指定されたファイルが違えば最適な翻訳環境も相違する場合が多くある、という点である。例えば、図4及び図5に示すように、"www.yahoo.com/"は「インターネット一般」が最適環境として登録されているが、同一WWWサーバ内にあるファイル"www.yahoo.com/Government/"や"www.yahoo.com/Recreation/Sports/"、"www.yahoo.com/Arts/"はそれぞれ「政治」及び「スポーツ」が最適環境として登録されている。更に、本発明者等の経験則に依れば、同一WWWサーバ内のファイルであっても、「パーシャルURL文字列」（すなわちURLを構成する文字列のうち"/"で区切られた各文字列）の組合せが最も類似する（すなわち最尤な）ファイル同士が同じ最適環境を共有する傾向にある、ということが既に判っている。例えば"www.yahoo.com/Arts/Architecture"というURL文字列で指定されるファイルは、"www.yahoo.com/"よりもパーシャルURL文字列の一一致度の高い"www.yahoo.com/Arts/"と同じ「アート」を最適な環境としている。

【0075】URLデータベース133へのURL文字列の登録は、例えば機械翻訳システム100の設計者が予め行っていてもよい。また、システム100のユーザが基本操作部135を介して追記登録するようにしてもよい（後述）。

【0076】URL入力部131は、WWWブラウザ画面上で入力されたURL文字列を、翻訳プロキシ110経由で受け取って、URL照合部134に渡すようになっている。また、URL照合部134は、入力されたURL文字列をURLデータベース133上で照合して、最適な翻訳環境を見つけ出してこれを翻訳環境出力部132経由で翻訳エンジン120に通知するようになっている。

【0077】C. 翻訳環境の選択処理

URL照合部134が行う処理は、要するに、入力されたURL文字列に最も類似した（すなわち最尤な）URL文字列をURLデータベース133の中で検索し、見つけ出されたURL文字列に対応する翻訳環境を翻訳エンジン120に出力することである。図6には、URL照合部134が行う翻訳環境の選択処理ルーチンをフローチャート化して示している。なお、該処理ルーチンにおいて、SはWWWブラウザ上で入力されたURL文字列、iはURLデータベース133上のインデックス、mは検索過程において今までに最も類似したURL文字列のインデックスを指す。例えば配列S(i)はURLデータベース133上のi番目のレコードに登録されたURL文字列を表す。また、データベース検索の最終結果

22

果がm=Mであれば、URLデータベース133上のM番目のURL文字列S(M)に対応する翻訳環境が最適であることを表す。また、配列E(m)は、URLデータベース133上のm番目のレコードに登録されたURL文字列の翻訳環境を示す。

【0078】該処理ルーチンでは、まず、各インデックスに初期値を代入する（ステップS10）。インデックスiの初期値は、URLデータベースの先頭レコードを示す1であり、インデックスmの初期値は、ブランク、すなわち類似のURL文字列が未だ見つかっていない状態を示す0である。

【0079】次いでステップS20では、現在のインデックスiがURLデータベース133の最後のレコードを越えているかどうかを判断する。

【0080】判断ブロックS20の結果が否定的、すなわち未照合のレコードがURLデータベース133に残っていれば、次ステップS30において、現在入力中のURL文字列Sがi番目のレコードに係るURL文字列S(i)をパーシャルURL文字列として含むかどうかを判断する（ステップS30）。

【0081】判断ブロックS30の結果が否定的、すなわちURL文字列S(i)が現在入力中のURL文字列Sに類似していないと判断された場合、ステップS70にジャンプする。ステップS70では、URLデータベース133の次のレコードを照合すべく、iを1だけインクリメントして、ステップS20に復帰する。

【0082】一方、判断ブロックS30の結果が肯定的、すなわちURL文字列S(i)が現在入力中のURL文字列Sに類似すると判断された場合、次ステップS40に進み、m=0すなわち未だ類似のURL文字列が見つかっていない状態かどうかを判断する。m=0であれば、この度照合されたURL文字列を一時記憶すべく、mにiを代入する。

【0083】また、判断ブロックS40においてmが0でない、すなわち類似のURL文字列が既に一時記憶されている状態であれば、次ステップS50に進み、この度照合されたURL文字列S(i)と既に記憶されているURL文字列S(m)のいずれが現在入力中のURL文字列Sに類似しているかを比較する。該比較は、文字列S(i)とS(m)のどちらが連続したより多くのパーシャルURL文字列を含むかを比べることによって行われる。例えば、"www.yahoo.com/"と"www.yahoo.com/Arts/"であれば、前者は1個、後者は2個のパーシャルURL文字列で構成されるから、後者の方が優る。なお、「連続したより多くのパーシャルURL文字列」とは、より多くのパーシャルURL文字列が一致していてもパーシャルURL文字列の並びが一致しなければ、URL照合上は類似性が低いことを意味する。例えば現在入力中のURL文字列が"www.yahoo.com/Recreation/Sports/"で、URLデータベース133に既登録のURL

50

23

文字列として"www.yahoo.com/Sports/"及び"www.yahoo.com/Recreation/"の2個があったとすると、後者は連続する2個のバーサルURL文字列が一致するが、前者は1個のバーサルURL文字列しか一致しない("Sports"は入力URL文字列中にあるが、並びが一致しないのでカウントされない)。したがって、後者のURL文字列の方が優る、ということになる。

【0084】判断ブロックS50において、新たに照合されたURL文字列S(i)の方が優ると判断されれば、次ステップS60に進み、最尤のURL文字列を置き換えるべく、mにiが代入される。しかし後、次ステップS70に進む。また、判断ブロックS50において、既に記憶されているURL文字列S(m)の方が優ると判断されれば、現在のmの値を維持すべく、次ステップS60をスキップしてステップS70に向かう。

【0085】ステップS70では、URLデータベース133の次のレコードを照合すべく、iを1だけインクリメントして、ステップS20に復帰する。そして、ステップS20において、未照合のレコードがURLデータベース133に残っていれば、上述のステップS30～S70の処理を繰り返す。一方、ステップS20で、現在のインデックスiがURLデータベース133の最後のレコードを越えていると判断されると、分岐"Yess"に進んでステップS80に向かう。

【0086】ステップS80では、まず、mが0かどうかを調べる。m=0は上記ループS20～S70において現在入力中のURL文字列Sに類似のURL文字列がURLデータベース133の中から全く見つからなかつたことを意味する。この場合には、デフォルトの翻訳環境を翻訳エンジン120に出力し(ステップS90)、翻訳環境の選択処理ルーチンを終了する。

【0087】一方、m≠0であれば、上記ループS20～S70において最尤と判断されたURL文字列に関する翻訳環境E(m)を翻訳エンジン120に出力し(ステップS100)、翻訳環境の選択処理ルーチンを終了する。翻訳エンジン120は選択された翻訳環境E(m)を用いてダウンロード・ファイルを翻訳処理する。

【0088】D. URLデータベースの更新

URLデータベース133の内容が基本操作部135を介して追記登録される、という点は既に述べた。この項では、URLデータベース133の追記登録手順について簡単に説明する。

【0089】図8には、URLデータベース133の追記登録手順ルーチンをフローチャート化して示している。まず、WWWブラウザ画面上に訳文テキストを表示し、これを見て翻訳内容に満足か、すなわち最適な翻訳環境が選択されているかどうか、ユーザに判断を仰ぐ(ステップS200)。

【0090】判断ブロックS200の結果が否定的であ

れば、追加登録の必要がないので、後続のステップを全てスキップして、処理ルーチンを終了する。

【0091】一方、判断ブロックS200の結果が肯定的であれば、新しく入力されたURL文字列Sと最適な翻訳環境E(m)との関係をURLデータベース133に追加登録すべきかどうか、ユーザの判断を仰ぐ(ステップS210)。ここで、ユーザは、新しく入力されたURL文字列Sの今後の使用頻度等を元に、登録の是非を判断する。

【0092】判断ブロックS210の結果が肯定的であれば、URLデータベース133に新規登録して(ステップS220)、処理ルーチンを終了する。また、判断ブロックS210の結果が否定的であれば、次ステップS220をスキップして、処理ルーチンを終了する。

【0093】E. ブラウザ画面

ここで、本実施例に係る機械翻訳システム100を実装したWWWブラウザ150の画面を、図7に例示しておく。同図の画面は、原文テキストの直下の行に訳文テキストを列記した「対訳表示」モードであるが、WWWブラウザ画面上で「原文のみ」又「訳文のみ」などの他の表示モードに切替可能である。また、WWWブラウザ画面はディスプレイ22上のマルチウィンドウの1つのウインドウでもよい。

【0094】WWWブラウザ画面には、最上段より、「タイトル・バー」、「メニュー・バー」、「ツール・バー」が設けられている。タイトル・バーは起動画面のアプリケーション名を表示した領域であり、メニュー・バーは選択可能な処理コマンドを一覧表示した領域である(周知)。また、ツール・バーは、頻繁に利用する機能を素早く起動するためのボタンを並べたメニュー・バーの一種であり、「スマート・アイコン」ともいう(周知)。

【0095】ツール・バーの真下には、URL入力部が設けられている。ユーザは、URL文字列をURL入力部に入力することによって、所望のアクセス先のWWWページを指定することができる。同図では、URL名"www.yahoo.com/"が入力されている。

【0096】入力されたURL文字列は、翻訳プロキシ110に渡される。翻訳プロキシ110は、WWWブラウザ150の代理で、指定されたWWWページからファイルをダウンロードするとともに、ファイル中の原文テキストを翻訳エンジン120に一旦引き渡す。また、入力されたURL文字列は、翻訳プロキシ110経由で翻訳環境切替部130に渡される。翻訳環境切替部130は、受け取ったURL文字列を解釈して最適な翻訳環境を選択し、これを翻訳エンジン120に通知する。また、翻訳エンジン120は、選択された翻訳環境を利用して原文テキストを翻訳処理し、原文テキストと訳文テキストを翻訳プロキシ110経由でWWWブラウザ150に渡す。これら一連の処理はデスクトップのバックグ

ラウンドで行われる。

【0097】URL入力部の下方の、ブラウザ画面の大部分を占める矩形領域は、WWWブラウザ画面本体であり、ダウンロードされたWWWページの内容が表示されている。同図では、URL名によって指定された"www.yahoo.com/"ページの画面が表示されている。

【0098】WWWブラウザ画面本体の最上段には、現在の表示モードや、選択された翻訳環境を表示するためのインジケータ部が設けられている。同図では、表示モードとして設定された「対訳」モードが強調表示されており、したがって、ブラウザ画面内では原文テキストの直下の行に訳文テキストが列記されている。また、図4に示すように、URLデータベース133には"www.yahoo.com/"の最適環境は「インターネット一般」と登録されていることから、インジケータ部にはその旨が表示されている。なお、ユーザは、インジケータ部をマウス27でクリック操作することによって、表示モードや使用された翻訳環境を動的に切り替えることもできる。

【0099】F. 追補

以上、特定の実施例を参照しながら、本発明について詳解してきた。しかしながら、本発明の要旨を逸脱しない範囲で当業者が該実施例の修正や代用を成し得ることは自明である。例えば機械翻訳システムは、WWWブラウザのプロキシ機能を利用したタイプである必要はなく、WWWブラウザ画面上の原文をクリップボード経由で受け渡しされて翻訳処理するタイプでアッフもよい。すなわち、例示という形態で本発明を開示してきたのであり、限定的に解釈されるべきではない。本発明の要旨を判断するためには、冒頭に記載した特許請求の範囲の欄を参照すべきである。

【0100】

【発明の効果】以上詳記したように、本発明によれば、複数の翻訳環境を備えた翻訳システムを用いて、外国語テキスト（例えば英語テキスト）を母国語テキスト（例えば日本語、あるいは韓国語や中国語など）に翻訳することができる、優れた機械翻訳方法及び装置を提供することができる。

【0101】また、本発明によれば、外国語テキストを母国語テキストに翻訳する際に、処理対象の外国語テキストに適した翻訳環境を選択して翻訳処理する、優れた機械翻訳方法及び装置を提供することができる。

【0102】また、本発明によれば、複数の翻訳環境を備えた翻訳システムを用いて、インターネット上のWWWサーバによって提供されたWWWページ中の外国語テ

キストを母国語テキストに好適に翻訳することができ
る、優れた機械翻訳方法及び装置を提供することができ
る。

【図面の簡単な説明】

【図1】図1は、本発明を実現するのに適した典型的なコンピュータ・システム10のハードウェア構成を模式的に示した図である。

【図2】図2は、本発明によって提供される機械翻訳システムのシステム構成を模式的に示した図である。

【図3】図3は、翻訳環境切替部130の構成を模式的に示した図である。

【図4】図4は、URLデータベース133をWWWブラウザ画面上に表示した様子を示した図である。

【図5】図5は、URLデータベース133をWWWブラウザ画面上に表示した様子を示した図である（但し、図4の画面の続き）。

【図6】図6は、URL照合部134が行う翻訳環境の選択処理ルーチンをフローチャート化して示した図である。

【図7】図7は、本実施例に係る機械翻訳システム100を実装したWWWブラウザ150の画面を例示した図である。

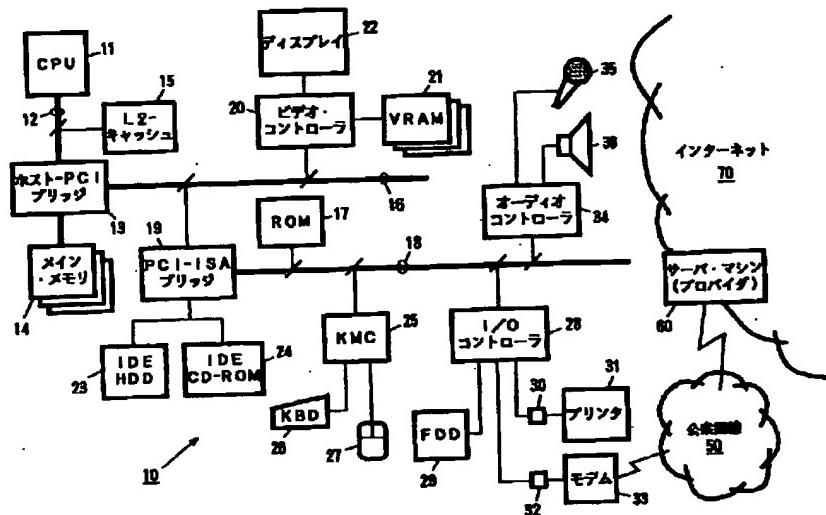
【図8】図8は、URLデータベース133の追記登録手続ルーチンをフローチャート化して示した図である。

【符号の説明】

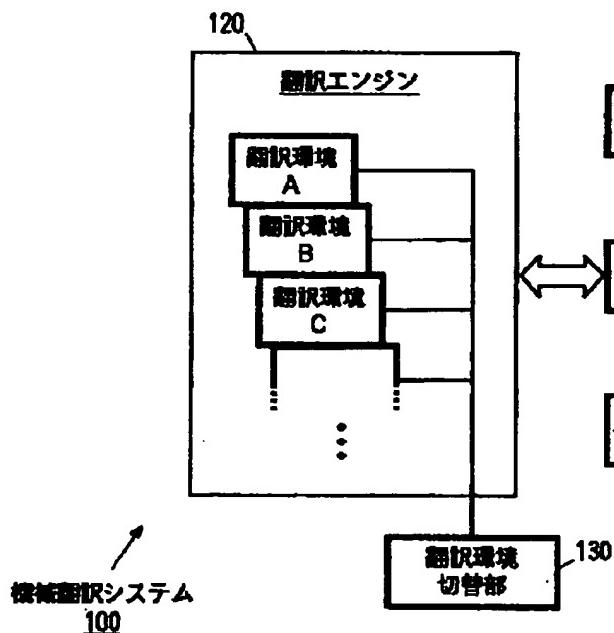
10…コンピュータ・システム、11…CPU、12…
プロセッサ・バス、13、19…ブリッジ回路、14…
メイン・メモリ、15…L2-キャッシュ、16…PC
Iバス、17…ROM、18…ISAバス、20…ビデ

30オ・コントローラ、21…VRAM、22…ディスプレ
イ、23…HDD、24…CD-ROMドライブ、25
…キーボード/マウス・コントローラ、26…キーボー
ド、27…マウス、28…I/Oコントローラ、29…
FDD、30…パラレル・ポート、31…プリンタ、3
2…シリアル・ポート、33…モデム、34…オーディ
オ・コントローラ、35…マイク、36…スピーカ、5
0…公衆回線、60…サーバ・マシン（プロバイダ）、
70…インターネット、100…機械翻訳システム、1
10…翻訳プロキシ、120…翻訳エンジン、130…
40翻訳環境切替部、131…URL入力部、132…翻訳
環境出力部、133…URLデータベース、134…U
RL照合部、135…基本操作部、150…WWWブラ
ウザ。

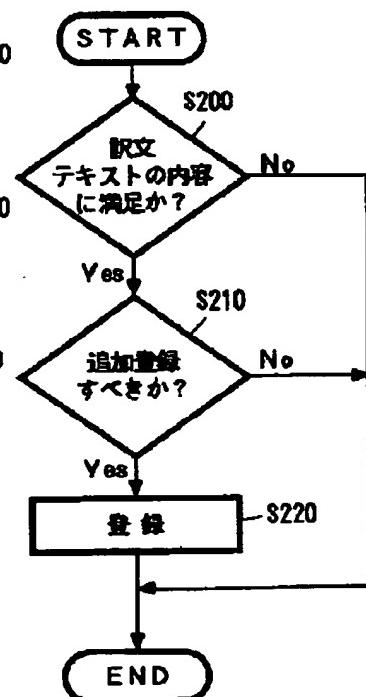
【図1】



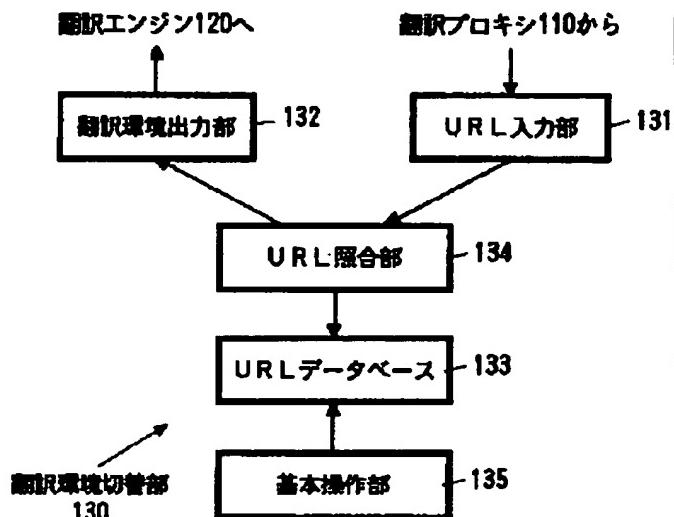
【図2】



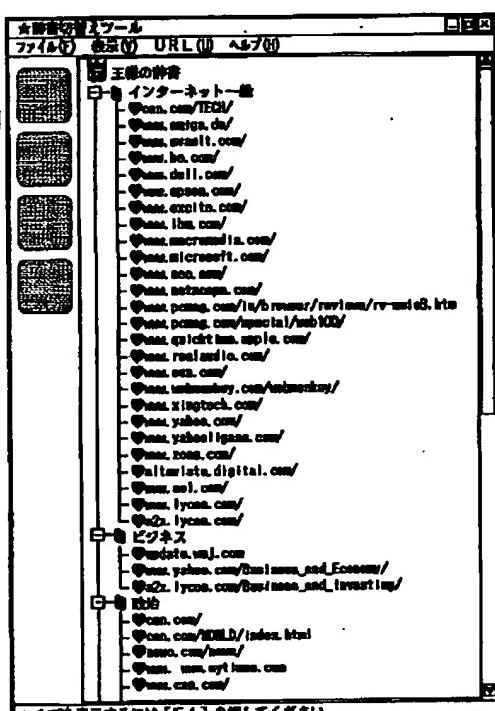
【図8】



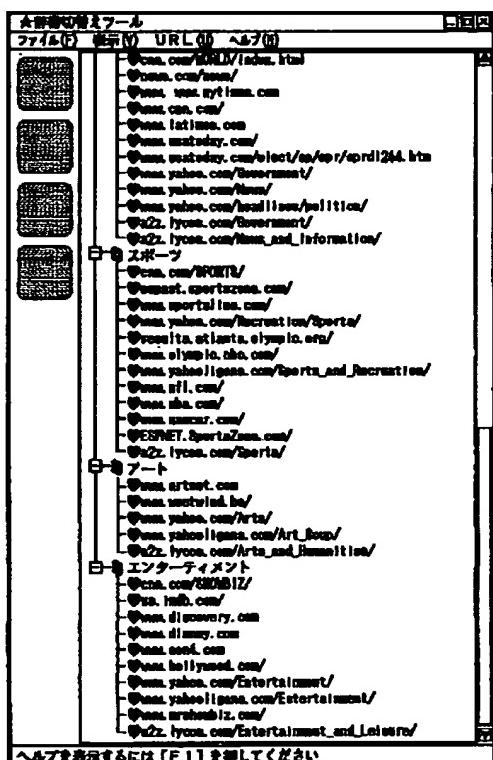
【図3】



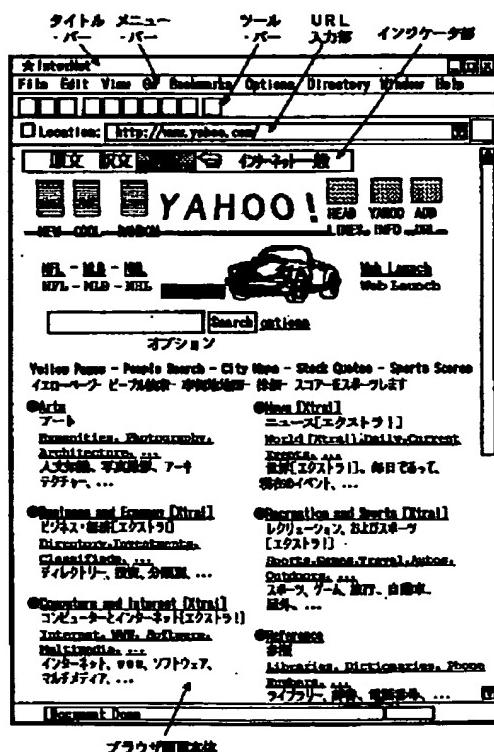
【図4】



【図5】

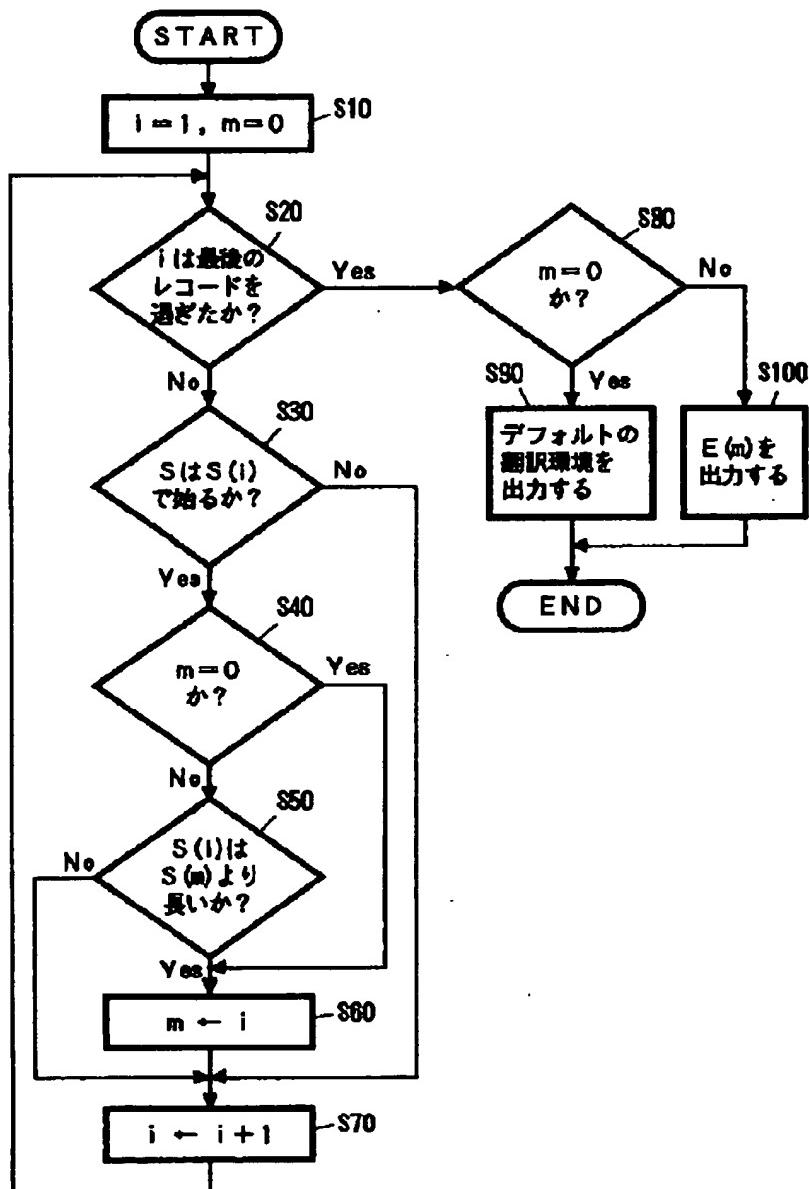


【図7】



ブラウザ画面表示

【図6】



フロントページの続き

(72)発明者 田添 英一
 神奈川県大和市下鶴間1623番地14 日本ア
 イ・ビー・エム株式会社 大和事業所内

(72)発明者 石倉 純
 神奈川県大和市下鶴間1623番地14 日本ア
 イ・ビー・エム株式会社 大和事業所内